DOCUMENT RESUME

ED 278 326 HE 020 009

AUTHOR Curry, Wade; Cruser, Paul
TITLE Can Testing Determine What Students Learn in General Education?
PUB DATE [86]
NOTE 15p.; Report on research conducted by Henry Wang.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Freshmen; College Seniors; College Students; Comparative Analysis; *Educational Assessment; Educational Objectives; *Educational Testing; *General Education; Higher Education; *Outcomes of Education; *Program Evaluation; Scores; *State Colleges
IDENTIFIERS *Trenton State College NJ

ABSTRACT

Testing to assess student learning in general education and the extent to which students achieve general education goals at Trenton State College were examined. Faculty reported that the general education program most emphasized both problem solving and communication and research and least emphasized the arts and non-Western civilization. Two tests that could be used to measure outcomes in general education were located: the College Outcomes Measurement Program (COMP) from the American College Testing Program, and Tests of General Education (TGE) from the Educational Testing Service. Seniors' tests results were compared with national norms and scores of transfer seniors and freshmen. Scores during the freshman and senior years were also compared, as were scores on the TGE and the COMP. In addition to summarizing findings of these comparisons, nine procedural problems in a general education testing program are identified. A six-step feedback loop for evaluating general education is suggested: developing goals, attaching the goals to each course, determining a means of assessment, testing and analyzing results, evaluating the instrument, evaluating and revising the curriculum, and revising the goals. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278326

CAN TESTING DETERMINE WHAT STUDENTS LEARN IN GENERAL EDUCATION?

by

Wade Curry and Paul Cruser

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Wade Curry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE 020 059

Can Testing Determine What Students Learn in General Education?

by

Wade Curry and Paul Cruser

Assessment in New Jersey

For the past 12 years, New Jersey has been testing all college freshmen for basic skills and providing enriched funding for remedial courses. The state Board of Higher Education has now mandated the testing of outcomes and established a committee to carry it out. That committee has rejected the "Florida Plan," in which the test is the only gateway into the Junior year. Such a test, they feel, would necessarily force colleges to concentrate on a narrow set of skills and on the very lowest standards. Instead, New Jersey, the committee recommends, should focus on improving academic programs through four kinds of assessment:

1. The final exams for remedial courses will be similar to the original basic skills test.

2. A state-wide test will focus on thinking skills.

3. Each college will assess the knowledge areas of its general education program.

4. Each college will assess the effectiveness of its major programs.

Wade Curry is the Dean of Arts and Sciences and Paul Cruser Assistant Dean of Arts and Sciences at Trenton State College. They report on research conducted by Henry Wang. Elizabeth Hager contributed significantly to their analysis and report.

Faculty Survey

At Trenton State College, a general education committee carried out a pilot assessment of our program. They were charged to determine whether and to what extent students achieved the college's eight general education goals. After revising the goals, the committee surveyed each department that offered general education courses, asking to what extent their courses met each goal and how they evaluated or could evaluate student achievement of these goals. Each member of the committee met with one or two departments to indicate what information was desired and how it would be used. Faculty were invited to appear before the committee to clarify their responses.

There was considerable resistance. Faculty assumed that the information would be used in personnel decisions or in determining which courses would remain in the general education distribution. Committee members tried to reassure the faculty that their purpose was only to get information about the current state of the curriculum.

The results were very interesting, after the committee clarified that there was no advantage in reporting that a course met all eight goals and that "moderate" emphasis did not mean a few words in a lecture. Several courses did not stress the goal that seemed their natural responsibility. For example, most humanities courses did not stress value clarification, and some stressed

met all eight goals and that "moderate" emphasis did not mean a few words in a lecture. Several courses did not stress the goal that seemed their natural responsibility. For example, most humanities courses did not stress value clarification, and some stressed neither Western nor Eastern culture. In courses for which more than one faculty member responded, there were sometimes very different profiles. It seemed to the committee that faculty should have the greatest freedom to choose materials, examples, theories, and teaching methods, but that they should agree upon the goals of a course.

The results of the survey are summarized in Table I.

TABLE I

GOAL	AVERAGE*
1. Communication and Research	2.10
2. Problem Solving	2.10
3. Value Clarification	1.44
4. Social Sciences	1.33
5. Mathematics, Science, and Technology	1.34
6. Arts	0.86
7. Western Civilization	1.26
8. Non-Western Civilization	0.83

* Very Strong = 3, Strong = 2, Moderate = 1, Not at all = 0

The faculty reported that our general education program emphasizes communication and research and problem solving; the least emphasis, they reported, was on the arts and Non-Western civilization.

Testing of Students

A second thrust of the committee was to assess student learning. Two of the faculty on the committee argued vigorously against any testing program. They said:

1. About half of our seniors are transfer students who took their general education in other colleges.

2. Our general education is a loose distribution of over 100 courses from which students or their departments select 15 courses. How can you test a distribution?

3. The gains made over the four years may result from variables other than classroom instruction. The less motivated freshmen will drop out, while the more motivated learners who remain will read widely and use their leisure for activities which develop their intellectual abilities.

4. The tests may simply measure intelligence. If so, they are a better measure of what the students bring to the college than what the college's program contributes to them.

These are all good arguments. We agreed with them and said, "Let's do it anyway." Finally, the majority of the committee agreed to permit us to seek a grant to run a pilot test.

Now that the pilot has been run, we must not only agree with the four arguments but add five more:

5. If the test doesn't affect the students' GPA or graduation, some will not try hard. A few of the essays were gibberish, although most students did seem to try hard.

6. The four hour test came at the end of a twelve hour day for most students. By ten o'clock, they were tired and irritable.

7. It is difficult to get people to give up four hours when there is no personal gain. We required the test, but a typical response was, "What will happen to me if I don't?" Only 70% of the students in our sample took the test, biasing the sample. Many of the better seniors did not take the test.

8. We could find no test that evaluates all the skills and understandings which we consider important. (It is probably impossible to construct such a test. One solution is that of Alverno College, which conducts many assessments, using a variety of methods.)

9. We were unprepared for the small differences in raw scores between freshmen and seniors. On the College Outcomes Measurement Program, the average difference is eight points on the total and two on each subtest category. At mid-range in a subtest category, a two point difference in raw score moves the student about 15 percentiles.

These nine procedural problems are extremely difficult to solve. We would suggest that enthusiasm be tempered by intelligence and skepticism, that results be interpreted with caution, and that several forms of assessment be used.

In any assessment program, the first step is to gain agreement on goals. The next step is to compare your goals to the tests available. As Table II indicates, you will not find many choices. (INSERT TABLE II.)

We were able to find only two tests designed to assess general education: COMP from American College Testing, and TGE from Educational Testing Service. What is worse, the TGE has not been normed

TSC Goals:

	GRE	MILL ANAL	NTE	CLEP	UAP	COMP	TGE
1. Comm., Rsch.	X	O	X	X	O	X	X
2. Prob. Solving	X	X	O	O	X	X	X
3. Clarify Values	O	O	O	O	O	X	O
4. Social Sciences	O	½X	X	X	X	X	O
5. Math, Sci., Tech.	½X	½X	X	X	½X	½X	X
6. Arts	O	½X	X	X	X	½X	O
7. Western Civ.	O	O	X	X	X	O	O
8. Non-western Civ.	O	O	O	O	O	O	O
Sub-Scores	X	O	O	X	X	X	X
Validated	X	X	X	X	X	X	O
National Norms	X	X	X	X	X	X	O
Essay	O	O	X	X	O	O*	X
Time	210	50	240	495	120	100	160
	GRE	MILL ANAL	NTE	CLEP	UAP	COMP	TGE

*** Also available in 240-minute Essay Form**

TABLE II

and is still considered experimental. Several tests were eliminated from consideration because they did not provide subtest scores; we felt that a single test score did not tell us much. The area examination from CLEP has some appeal because it covers knowledge areas broadly, but it is far too long. The Graduate Record Examination has a number of supporters, but it seems to us a poor choice. It is intended to discriminate at the upper end of the range; for example, on the GRE Chemistry examination you get a perfect score if you get two-thirds of the questions correct. If we must choose, we would prefer that the test discriminate best at the lower end of the scale, so that we can be sure that our students graduate with certain intellectual skills. The Miller Analogy is a good test of intelligence and therefore useful as an admissions test but not as an assessment test. The general examination of the National Teachers Examination equates well with our goals and is a fairly good test (except for critical thinking); however, at the time we were considering it, ETS would not provide subscores. The Undergraduate Program Test, like the GRE, seems to be a selection test, not an outcomes measurement device.

We were left with the COMP and the TGE. The COMP is geared to testing how a person will function as an adult. Our aim as colleges is to help students function effectively as adults. But if you were to ask a sociology professor, "Are you teaching how to function effectively in a social institution?", he or she would say, "No, I'm teaching sociology." For that reason, COMP is not a very good measure of what is actually being taught. A second weakness is that with each question, the testers are measuring

three different goals. The questions cannot be too specific or too hard (for example, they cannot have any questions on classical music or on calculus), because they would be losing information in three different areas.

The Tests of General Education concentrate on four intellectual skills and do not measure content. They have not been validated or normed, and we do not know what ETS plans to do with them.

We decided that we wanted to make five comparisons:

1. Our native seniors with our transfer seniors.
2. Our native seniors with national norms.
3. Our native seniors with our freshmen.
4. Our freshmen with their scores as seniors (in three years).
5. We wanted to compare the TGE and the COMP.

Because our faculty were most concerned about weaknesses in writing, we gave the essay examination from TGE to students taking the COMP as well as to those taking the objective TGE.

On the COMP, the difference between our sample of seniors and the sample of freshmen was eight raw score points, about average, but not enough to give us a lot of confidence in the results. In percentiles, the way ACT and individual colleges always present it, the gain seems huge; in actuality, it is not.

(INSERT TABLE III.)

Table III compares our students with freshman national norms. What you see is not what the faculty expected. They remain unconvinced. "If <u>this</u> writing is at the 60th percentile, what must it be like to teach in the rest of the country?" But communication

COMP: Mean Percentiles for TSC Freshmen on National Norms

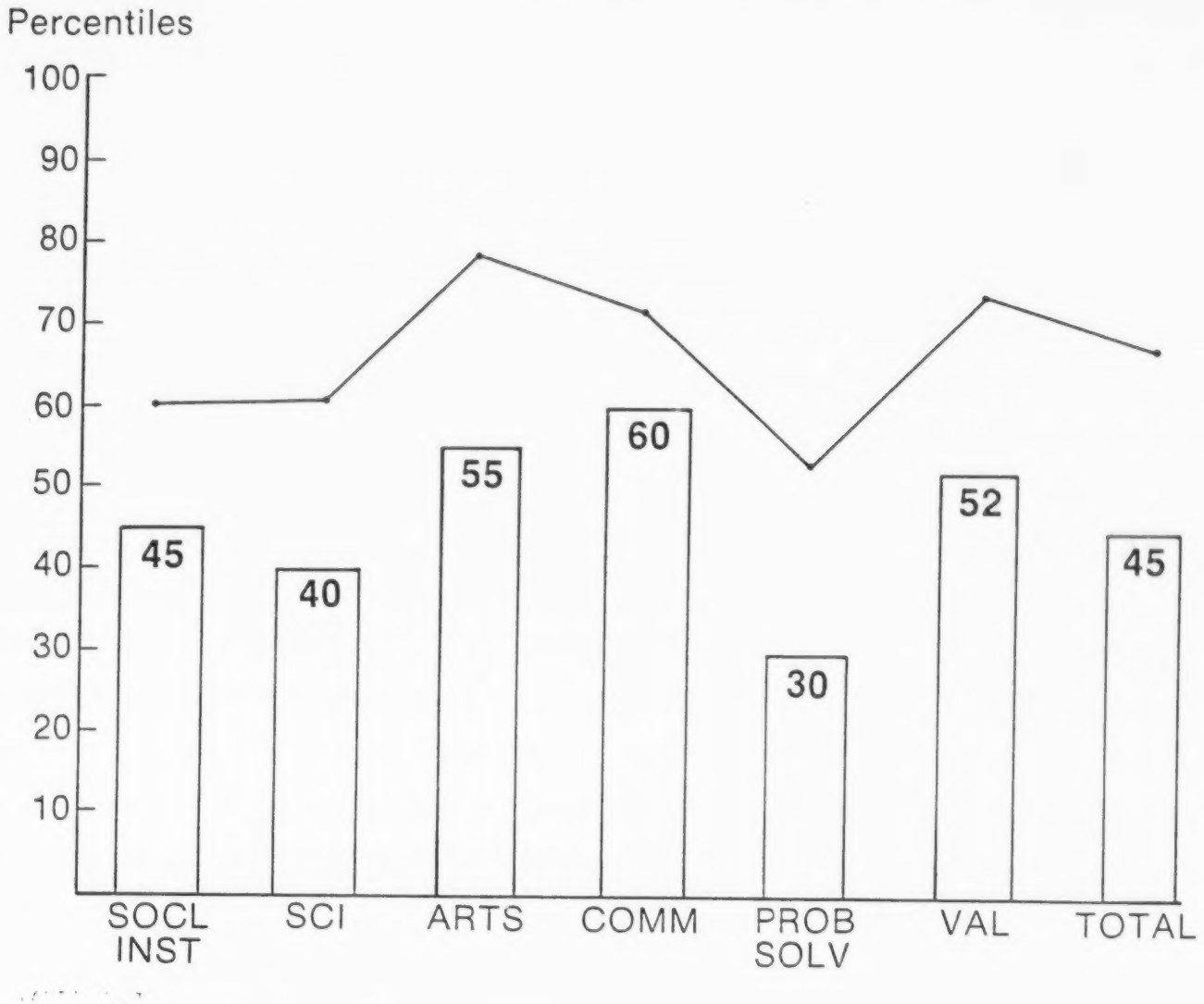

TABLE III

appears to be an area of strength for our students. Faculty had reported that they emphasized least the areas of value-clarification and the arts; yet, those are areas of strength for our students. The most obvious valley on the chart is problem-solving, a goal which our faculty say they strongly emphasize. The line graph of senior scores is a mountain range that has the same peaks and valleys as that of the freshmen. The students go out looking pretty much like they did when they came in. They don't remedy their weaknesses. Why should we expect them to? The elective curriculum permits them to select courses in which they already have strength. College is designed to strengthen strengths and to avoid courses in which weaknesses will be revealed. Is that what we intended?

There are no national norms for the Tests of General Education. We compared our students to those at a state university. Their weakest thinking skill was integration, but they did better than the university students on the other three skills. This weakness makes sense. Our general education program is a series of discrete introductions to disciplines. Few courses have hooks out to other disciplines, nor do they require students to integrate what they have learned. They should.

(INSERT TABLE IV.)

Our biggest surprise was that the difference between our men and our women was greater than the difference between our freshmen and our seniors. As Table IV shows,females far outscored males in communication, and males far outscored females in mathematics. They come in that way and they go out that way. Apparently, the

TGE: Male/Female Senior Comparison Reveals Difference in Skills

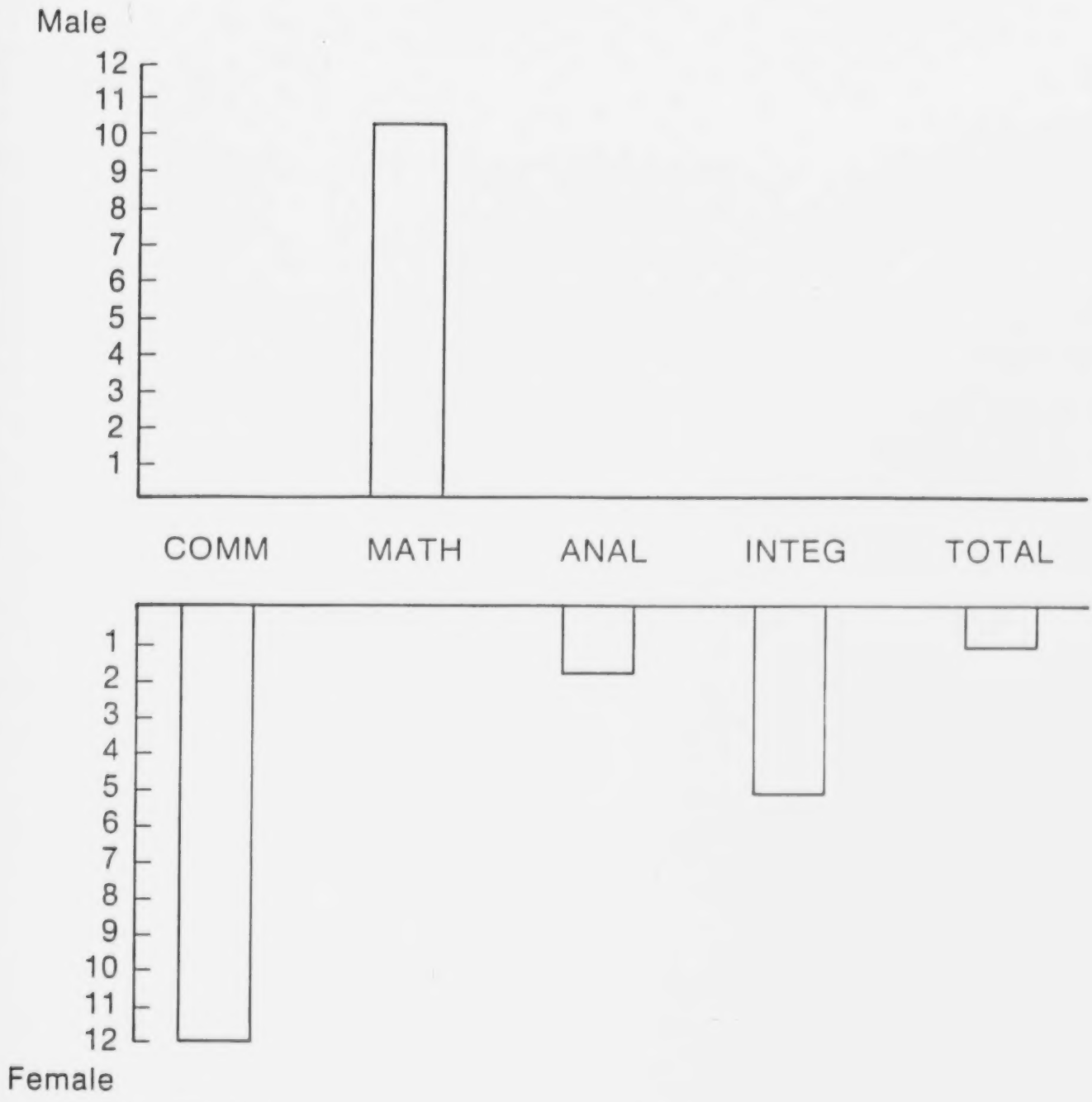

TABLE IV

average woman is not going to take a course that requires quantitative work, and the average man is not going to take a course that requires writing. They perceive not a need to remediate a weakness, but a need to avoid revealing the weakness.

On the eight essays, our seniors usually scored far higher than our freshmen. Our major concern, however, was that about 20 to 30% of the seniors scored low on each question. While this group was not always the same students, there was enough consistency to make us feel that a significant number of seniors (perhaps 10% to 15%) graduate without good communication skills and without cultural awareness. I think that we should be more concerned about those students than about whether our average student is above or below national norms.

There was no significant difference between our native and our transfer seniors. That surprised us since we are very selective and the community colleges in our state are open-door.

Conclusions

Are the COMP and the TGE good tests? The total score on the COMP revealed a good (but not an outstanding) correlation with the grade point average (.36) and the SAT (.43) score. The correlations of the six sub-scores with the total score were very high, as you would expect since each question measures three skills. The correlations of the TGE with the GPA (.47) and the SAT (.61) are considerably higher than those of the COMP, and the correlations of each test with the total score are also high. Of the two tests, we prefer the TGE because it seems to to do a better job of isolating the intellectual skills and because the raw score

ranges are greater. Some of the distinctions made by COMP seem to us a little dubious. For example, the higher order communication skills branch over into critical thinking and into artistic expression. How could our students score so high on communication and so low on problem solving when the two areas of skill are so closely related?

Is assessment worth doing, when you consider all the problems in getting a good test and a good sample of students? Our experience would seem to provide a resounding Yes. After all, we found that the students did not demonstrate the skills that the faculty said they were stressing or the weakness they expected. The assessment also provided clear evidence that our courses will have to be delivered in such a fashion that students will be required to develop the intellectual skills in order to pass them.

The assessment process to us is not a way of evaluating the individual student or the individual professor, but a way of evaluating our program. To us, the process is a six-step feedback loop.

1. Develop or revise goals.
2. Attach the goals to each general education course.
3. Select or create a means of assessment.
4. Administer it and analyze the results.
5. Evaluate the instrument.
6. Evaluate and revise the curriculum.
1. Revise the goals

Continue...